Draft Lab Book Notes

Analyst	Maria Battle
Project	Preventive Maintenance Work Order Performance
Original dataset	FMDataset1-25-2019-share.xlsx
Primary / Cleaned Dataset	FMDataset_copy1.csv
Data Sources	https://github.com/MariaBattle/Baltimore-Dept-of-General-Services, https://data.world/euclid46/analysis-of-baltimore-city-preventive-maintenance, https://data.world/brl1906/facility-maintenance-data-jan-2019, https://data.baltimorecity.gov/resource/689q-eiea.csv, https://data.baltimorecity.gov/d/g5wn-pq7y.csv https://www.ce.jhu.edu/baltimorestructures/Index.php?location=Clarence%20M.%20Mitchell,%20Jr.%20Courthouse
Code Files	https://github.com/MariaBattle/Baltimore-Dept-of-General-Services/blob/master/APIs

General Description

The primary dataset contains 87,278 observations. Data has been sanitized by the DGS legal department which may have had PII and system part numbers that were internally cataloged. The dataset was downloaded on January 30, 2019 from https://data.world/brl1906/facility-maintenance-data-jan-2019.

Lab Book Index

1. Data Element Dictionary (DED)
2. Values Description
3. Data Quality
4. Data Cleaning
5. Data Transformation
6. Exploratory Data Analysis Ideas
7. New Variables

1. **Data Element Dictionary (DED) of Original Dataset -** FMDataset1-25-2019-share.xlsx

Column in Original Dataset	Column Header	Description
A	act_labor_hours	Actual labor hours spend on a work order. DGS has ~30% confidence that these labor hours are accurate.
B	activity_type	
C	allow_work_on_holidays	Whether work was allowed to be conducted on a holiday to fulfill the work order
D	bl_id	building identifier
E	cause_type	tag to capture instances where corrective maintenance was found in the course of doing HVAC preventative maintenance work.
F	cost_est_labor	Estimated cost of work order labor
G	cost_est_other	Estimated cost of work order other than for labor
H	cost_est_parts	Estimated cost of parts used to fulfill work order
I	cost_est_tools	Estimated cost of tools used to fulfill work order
J	cost_est_total	Estimated total cost of labor, other, parts, and tools
K	cost_labor	Actual cost of work order labor
L	cost_other	Actual cost of work order other than for labor
M	cost_parts	Actual cost of work order parts used to fulfill work order
N	cost_tools	Actual cost of work order tools used to fulfill work order
O	cost_total	Actual total cost of labor, other, parts, and tools
P	curr_meter_val	
Q	date_assigned	date work request was assigned to a team or technician to address as a valid work request
R	date_closed	date all work and payment associated with a work order is determined to be done.
S	date_completed	date the required maintenance or repair work is completed.
T	date_escalation_completion	
U	date_escalation_response	
V	date_est_completion	Estimated data that the work order was completed/closed.
W	date_requested	Date work order was requested by a service request liaison for a building in the portfolio or requested by an internal staff maintenance person.
X	date_stat_chg	
Y	desc_other_costs	Description of other incurred costs that were not captured/anticipated

Z	dp_id	Numerical department ID that corresponds to dv_id character string code designation for each city department
AA	dv_id	Character string code designation for each city department that requested a work order.
AB	eq_id	
AC	escalated_completion	
AD	escalated_response	
AE	est_labor_hours	Number of hours used to fulfill the work order
AF	option1	
AG	option2	
AH	pmp_id	Preventive type preventive maintenance; string value of the test or inspection type.
AI	pms_id	Non-unique preventive maintenance ID number corresponding to type of system.
AJ	prob_type	Work request problem type. Captures the primary category of the type of maintenance request and determines generally who is assigned to complete the work.
AK	repair_type	Type of repair applied to the work order
AL	res_id	
AM	rmres_id	
AN	rsres_id	
AO	satisfaction	Customer satisfaction scale
AP	site_id	
AQ	status	
AR	time_assigned	Eastern standard time the work order was assigned to a person as a task (not the time it was entered into the system). Corresponds to date columns.
AS	time_completed	Eastern standard time the work order was completed and closed.
AT	time_escalation_completion	
AU	time_escalation_response	
AV	time_requested	The time the work order was requested and entered into the system.
AW	time_stat_chg	
AX	tr_id	
AY	work_team_id	The repair team assigned to the work order
AZ	status_sort	
BA	not_to_exceed	Work order budget cap, a dollar amount not to be exceeded by the work order
BB	name	Building name or physical address
BC	time_start	

BD	time_end	

2. **Values Description -** FMDataset1-25-2019-share.xlsx

Column Header	Values Description
dv_id	B.O.E – Board of estimates BOE - Board of estimates BOESTIMATES - Board of estimates BCPD – Baltimore City Police Department POL – Baltimore City Police Department POLICE – Baltimore City Police Department POLICE DEPT – Baltimore City Police Department COMPTROLLER – Comptroller Office COMPTROLLERS – Comptroller Office D.G.S – Department of General Services DGS - Department of General Services D.O.T – Department of Transportation DOT – Department of Transportation DHCD – Department of Housing and Community Development H.C.D - Department of Housing and Community Development HCD - Department of Housing and Community Development FIRE – Fire Department FIRE DEPT. – Fire Department M.O.I.T – Mayors Office of Information Technology MOIT - Mayors Office of Information Technology PARK – Parks and Recreation RECPARKS – Parks and Recreation PARK. AUTHORITY – Parking Authority PARKING – Parking Authority PLAN – Department of Planning PLANNING – Department of Planning

	BBMR – Bureau of Budge and Mgmt Research BCFD – Baltimore City Fire Department BCIT – Baltimore City Information & Technology BOPA – Baltimore Promotion & the Arts CITY COUNCIL –City Council CITY HALL - City Hall COURTS – courts D.P.W – Department of Public Works DOF – Disability Opportunity Fund DPW – Department of Public Works FIN – Department of Financial Assistance FINANCE – Department of Finance HABC – Housing Authority of Baltimore City HEALTH – Health Department LIBRARIES - libraries MOED – Mayor's Office of Employment Development MOHS – Mayor's Office of Human Services PABC – Pedestrian & Bicycle Advisory Committee PERSONNEL – Human Resources YWCA - YMCA dp_id and dv_id are departments and divisions. The divisions are within the department and hence some correlation.
activity_type	SERVICE DESK – MAINTENANCE – the request was received by the maintenance department service desk & entered into the system.
allow_work_on_holidays	1 = allowed (yes) 0 = not allowed (no)
bl_id	Codes don't need to be described as long as they are associated with a dv_id. Please see Data Quality Notes.
cause_type	CM from PM = corrective maintenance from preventive maintenance Blank = there was no correction. For example: if a technician goes for a PM and realizes that there is more work to be done, it no longer remains a PM, but turns into a CM.
cost_est_labor	Numerical values are in one dollar units
cost_est_other	Except for bl_id B00027, all values are blank (0).
cost_est_parts	Except for bl_id B00100, all values are blank (0)
cost_est_tools	All values are blank (0)
cost_est_total	Numerical values are one dollar units
cost_labor	Numerical values are one dollar units
cost_other	Numerical values are one dollar units

cost_parts	Numerical values are one dollar units. Negative values are not valid.
cost_tools	All values are blank (0)
cost_total	Numerical values are one dollar units.
curr_meter_val	All values are blank (0)
date_assigned	Date the work order was assigned to a person as a task; time fields are blank; 19,824 observations are blank for this variable.
date_closed	Date and time the work order was closed and the maintenance issue was resolved. Closed = when the work order is completely done including invoicing. The work order is then funneled to historical bank ~13,800 observations are blank for this variable. Blank means that the work-order was completed but not closed.
date_completed	Date the work order was completed. Completed: It occurs before close. The physical work is done but other portions after service may be pending.
date_escalation_completion	All values are blank (0)
date_escalation_response	All values are blank (0)
date_est_completion	All values are blank (0)
date_requested	Date the work order was requested and entered into the system.
date_stat_chg	All values are blank (0)
desc_other_costs	All values are blank (0)
dp_id	Mix of character and numerical. dp_id and dv_id are departments and divisions. The divisions are within the department and hence some correlation.
eq_id	Numeric values.The numerical value is tagged to the equipment by the vendor. It is just a number given to the equipment. (Anything that is numbered is HVAC units)
escalated_completion	All values are blank (0)
escalated_response	All values are blank (0)
est_labor_hours	Float values. Scale is hours
option1	All values are blank (0)
option2	All values are blank (0)
pmp_id	GENERATOR PM = generator preventive maintenance inspection FUEL TANK TEST= fuel tank test for preventive maintenance inspection BLDG INSPECTION = building inspection. This category is for everything that does not have designated inspections such as general items, windows doors, paints, ceilings, non system items, non HVAC, Non generator, non elevator, non fuel tanks are included in building inspection. HVAC INSPECTION = heating, ventilation, A/C inspection POOL INSPECTION = swimming pool preventive maintenance inspection BASEMENT INSPECT = basement preventive maintenance inspection ELEVATOR TEST = elevator preventive maintenance inspection

	EXTERMINATION = extermination preventive maintenance inspection FILTER CHANGE = filter was changed as a result of preventive maintenance inspection FLOOR BUFFING = floor cleaning services HEAT WATER TEST = water heater preventive maintenance inspection HEATING LEVELS = heating system preventive maintenance inspection KITCHEN PM = kitchen preventive maintenance inspection PEST CONTROL = pest control preventive maintenance inspection UTIL RM B00064 = utility room preventive maintenance inspection If an observation does not have a pmp_id (i.e. blank), there is no PM/CM attached to that work. It is just a straightforward work order.
pms_id	Numeric values that do not have a meaning for the system or any part ofr the work order
prob_type	String description of problem type. There are 22,865 dv_id blanks for this variable.
repair_type	All values are blank (0) except for 3 observations
res_id	All values are blank (0)
rmres_id	All values are blank (0)
rsres_id	All values are blank (0)
satisfaction	Customer satisfaction scale. What does each value mean?) 0 = 1 = 2 = 5 =
site_id	Values are same as dv_id unless highlighted blue. A site is just where a group of buildings is housed. The site is not tied to the division information. (dv_id) B BCFD BCPD BCPSS BOSTONMARINE C CIVIC CENTERS D DGS DGS HISTORIC DOT DPW

<table>
<tr><td></td><td>DPW SOLID WASTE
DPW STILLEY
DPW W/WW
DPW WASTE WATER
DPW WATER SUPPLY
DPW-DAMS
E
ELECTIONS
FINANCE
HCD
HEALTH
LIBRARIES
MOED
MULTI
PARKING AUTH
PRIVATE
PUBLIC ASSEMBLY
RECPARKS
SERVICE CLASS C
TEST
UNLISTED
VISIT BALTIMORE</td></tr>
<tr><td>status</td><td>Status of work orders e; does the system or a person enter/update that status?<table><tr><td>A = Approve</td></tr><tr><td>AA = Assigned to Work Order</td></tr><tr><td>Can = Cancelled</td></tr><tr><td>Clo = Closed</td></tr><tr><td>Com = Completed</td></tr><tr><td>HA = On Hold for Access</td></tr><tr><td>HC = Referred to Contractor</td></tr><tr><td>HI= Waiting on Invoice</td></tr><tr><td>HL = On Hold for Quote</td></tr><tr><td>HMI = On Hold for More Information</td></tr><tr><td>HP = On Hold for Parts</td></tr><tr><td>I = Issued and In Process</td></tr><tr><td>R = Requested</td></tr><tr><td>Rej = Rejected</td></tr><tr><td>RMI = Request for More Information</td></tr><tr><td>S = Stopped</td></tr></table></td></tr>
</table>

time_assigned	Eastern standard time the work order was assigned to a person as a task (not the time it was entered into the system). Corresponds to date columns.
time_completed	Eastern standard time the work order was completed and closed.
time_escalation_completion	All values are blank (0)
time_escalation_response	All values are blank (0)
time_requested	Invalid dates: Jan 0, 1900. Time is Eastern Standard time zone. These values correlate to date_requested. Are the times correct? Are the times Eastern Standard time?
time_stat_chg	All values are blank (0)
tr_id	Type of work order (by category)
work_team_id	The name of the repair team who conducted the work.
status_sort	Data storage values for the system. These do not have meaning for the work orders.
not_to_exceed	Single dollar units. The quote is provided upfront and the amount in the column is the proposed and agreed amount to provide the service and the material.
name	Building name or physical address
time_start	The time that the work started on a given date
time_end	The time that the work ended on a given date

3. **Data Quality Notes**

Building and Division Identification Numbers. After removing duplicates in dv_id, there are 29 unique Division IDs (dv_id) with multiple building identifiers (bl_id). Of the ~87,000 observations, 22,865 observations are blank for these 2 variables. Assuming every data point must be associated with either / both of these variables, our dataset could be reduced to 64,135 observations; however, other variables are also blank and we'll need to determine how to handle those observations and variables.

Possible Duplicate Values. Please see Section 2, Values Description. Possible duplicates are highlighted in yellow.

Preventive Maintenance ID. pm_id values exist for the following dv_id only:

- BCPD – 44 pm_id observations
- City Hall – 1 pm_id observation
- DGS – 1,321 pm_id observations
- DOF – 25 pm_id observations
- DOT – 11 pm_id observations
- POL – 5 pm_id observations

- (blank dv_id) – 3,482 pm_id observations

Values in act_labor_hours. When correlated with cost, these hours possibly present a heavy skew. When correlated with prob_type, some observations show thousands of dollars charged for 0 hours, and others show a charge of $6,500 per hour.

4. **Data Cleaning Notes**

The cleaned version of dataset FMDataset1-25-2019-share.xlsx is named FMDataset_copy1 and is located here on data.world.

Column Header	Values
dv_id & dp_id columns	Duplicates were removed as the following: B.O.E, BOE, BOESTIMATES – normalized to BOE BCPD, POL, POLICE, POLICE DEPT - normalized to BCPD COMPTROLLER, COMPTROLLERS – normalized to COMPTROLLER D.G.S, DGS – normalized to DGS D.O.T, DOT – normalized to DOT DHCD, H.C.D., HCD – normalized to DHCD HCD/HABC converted to DHCD_HABC FIRE, FIRE DEPT., BCFD – normalized to BCFD M.O.I.T, MOIT – normalized to MOIT PARK, RECPARKS – normalized to RECPARKS PARK. AUTHORITY, PARKING – normalized to PARKING PLAN, PLANNING – normalized to PLANNING MOED, M.O.E.D - normalized to MOED FIN, FINANCE - normalized to FINANCE DPW, D.P.W. - normalized to DPW
cost_est_tools cost_tools curr_meter_val date_escalation_completion date_escalation_response date_est_completion date_stat_chg desc_other_costs escalated_completion escalated_response	Removed these columns due to no data

option1 option2 res_id rmres_id rsres_id time_escalation_completion time_escalation_response time_stat_chg	
name	Enoch Pratt Free Library, EPFL - normalized to EPFL Ineer Harbor Promenade corrected to Inner Harbor Promenade
Date_assigned date_requested	Removed time stamps because they were all zero
time_requested	Removed column because all dates were January 1900

5. **Data Transformation Notes**

6. **Exploratory Data Analysis Ideas**

Additional Datasets

An additional dataset has been provided: Latlong.xlsx (renamed to FMD_latlong.xlsx), however there were too little rows in the latlong sheet to join on FMDataset_copy1. Alternatively, the firedata dataset, located at https://data.baltimorecity.gov/resource/689q-eiea.csv, can be used to join on a subset of FMDataset_copy1 Fire Station observations in the `name` variable that contain engine, truck, medic, and EMS numbers. For example, E-2 in the firedata dataset corresponds to Engine 2 in the FMDataset_copy1 dataset. This correlation could provide addresses, police district and council district information for Fire Departments which have the highest labor hours.

The budget dataset located at "https://data.baltimorecity.gov/d/g5wn-pq7y" has budget information relating to dv_id (e.g., Fire) and expense_category (e.g. Facilities Maintenance and Replacement). This information appears to be more robust than the estimated cost columns in FMDataset_copy1 and provides expense_type categories. *Need to confirm data quality and assumptions with Babila

Possible Work Order Performance Indicators in FMDataset_copy1.csv

- Frequency of corrective work orders that indicate a system is dying correlated with age, usage,

PM

- Whether PM scheduling is fixed (must be performed at a specific time), variable (required, but with some flexibility in the time of performance), or optional (desired, but may be skipped without immediate consequences).
- Compare preventive and corrective maintenance due dates with date of ticket completion.
- Increasing / decreasing overtime labor cost (indicators of emergency work that could result from lack of upkeep)
- Increase/decrease in corrective work orders since preventive maintenance was implemented.
- Cost of labor exceeding the cost of replacement parts
- Estimated cost to address a building's physical needs over the next ten years
- Proportion of maint resources / activities that are purely reactive.
- Proportion of maint resources / activities that are dedicated to preventive maintenance
- Proportion of corrective versus preventive work orders
- Frequency of corrective work orders using groupings of other variables
- Cost/benefit between preventive and corrective
- Process efficiency of frequency and time duration of work orders (decreased time indicates time is freed to do other tasks)

How Data Quality and Data Capture Affect Results

...

...

...

How DGS Can Improve Data Quality and Data Capture

...

...

…

Initial Observations

Labor Hours by dv_id:

BCFD - 16,807 hours
(blank) - 8,127 hours, 227 hours are CM from PM, the most data for the cause_type variable
DGS - 3,643 hours
BCPD - 2,680 hours
DOT - 1,450 hours
DPW - 389 hours

MOHS - 324 hours
BOPA - 60 hours

**I need to clearly identify which observations are corrective versus preventive.

229.93 of 34,396 labor hours are categorized as CM from PM (cause_type variable)

Out of $28,325,097.77 spent on labor for maintenance, $28,283,284.69 was not categorized as preventive maintenance under pmp_id (meaning is was corrective maintenance?). The $41,813 difference was categorized as preventative maintenance in the pmp_id variable. Identify the year(s) this pm cost was spent and calculate the difference in maintenance work order rate and labor cost from years with mostly/only corrective maintenance.

A possible correlation: preventive work (pmp_id) and corrective maint work (prob_type) quantified by cost_labor. Sort by prob_type first, then pmp_id. Blank pmp_id observations may indicate corrective maintenance.

Develop subsets by fiscal year and corrective vs. preventive maintenance; track cost and # of days for open orders for categorical maintenance and dv_id; estimate trend over next 10 years.

Possibly Add the following new variables:

- Difference of days between date requested and date assigned.
- Difference of days between date assigned and date completed.
- Difference of days between date completed and date closed.
- Proportion of cost labor, parts, and other to the total cost. See data quality notes on act_labor_hours.
- Unless otherwise specified, join coordinates data of locations associated with highest values in the work order duration and cost proportions.
- Historical weather temperature correlation to frequency of work orders in certain years/prob_type.

7. **New Variables in Cleaned Dataset -** FMDataset_copy1.csv

year_built	https://www.ce.jhu.edu/baltimorestructures/Index.php?location=Clarence%20M.%20Mitchell,%20Jr.%20Courthouse Note: I will consult with a librarian for the rest of the building ages.
bldg_age	current year (2019) minus the year the building was originally built
time_dur	time_start minus time_end, scale is in hours
day_dur	Number of days between date requested and date completed

tkt_freq	Grouped by `name`, the number of days between corrective service requests for each building
over_maint	The dollar amount difference between cost of labor and cost of parts. In some cases, if maintenance labor costs more than parts, this system may be considered for replacement or run-to-failure.